



Vermont Program for Quality in Health Care, Inc.

PHASE III REPORT

Contract # 39548

**DEVELOPMENT OF SPECIFICATIONS FOR SET OF ANALYSIS-
READY HEALTH CARE DATA PRODUCTS**

January 27, 2021

Prepared by:

Catherine E. Fulton, Executive Director, VPQHC

Hillary Wolfley, Health Data Analyst, VPQHC

Steve Kappel, Policy Integrity LLC

David Healy, Consultant, Stone Environmental

Contents

Executive Summary.....	3
I. Phase I and II Overview.....	4
II. Phase III Introduction and Overview.....	6
Chapter 1 – General Principles and Considerations	7
Chapter 2 – High Level Framework.....	8
Introduction	8
Design Data Structures Based on the Operation of Underlying Process, Rather Than on the Format of Input Documents	9
Maximizing the Amount of Knowledge Built into the Data Structures	9
Make Records Meaningful To Count	10
Separate Financial and Clinical Information	10
Simplify Timing Challenges Inherent in the Data.....	12
Provider and Individual Identifiers.....	13
Support Files	14
Related Records to Place	15
Improve Data Quality / Trust in Data.....	18
Weighting.....	18
Other Data	19
Chapter 3 - Methods for Meaningful Public Use Files	20
Introduction	20
Chapter 4 - Recommendations to Address Different User Types.....	22
Introduction	22
Chapter 5 – Recommendations Summary	25
Appendix 1 – Summarization of Claims into Clinical Events	28
Appendix 2 – Enrollment Report	29
Analysis of Commercial Enrollment Data in VHCURES	29
Appendix 3 – Possible Data Model	35
Appendix 4 – Acronyms	36

Executive Summary

The Green Mountain Care Board (GMCB) and its Analytics Team have undertaken a process to improve access, optimize functioning, and increase utilization of the Vermont Healthcare Claims Uniform Reporting and Evaluation System (VHCURES) data set for ongoing analysis. The Phase III Development of Specifications for Set of Analysis-Ready Health Care Data Products report that follows is the culmination of a process to evaluate opportunities and available mechanisms to produce analytic files that will support greater and easier access to VHCURES data.

The Phase I report contained a review of previous efforts to redesign VHCURES within the past few years, including recommendations made by Bailit Healthcare Consulting and Truven Analytics. Summary information from prior efforts helped inform the direction and scope of continuing evaluation activities under this work.

Phase II activities included stakeholder interviews that identified specific needs and feedback that further defined priorities for redesign that would encourage new user utilization. A summary table of “themes” that emerged from the stakeholder interviews is included in the Introduction section of this report. Themes included:

- Looking at Total Cost of Care per Capita by Age, Gender, and Health Service Area (HSA).
- Comparison of Utilization between ACO-attributed and non-attributed populations and across time.
- Spending on Patients with Specified Chronic Illnesses Over Time.
- Integrating Information from Birth and Death Certificates with VHCURES.
- Match VHCURES Inpatient Facility Claims with VUHDDS.

Phase II also included a review of All Payer-Claims Database (APCD) development in all 50 states to ascertain emerging best practice and continuing progress in APCD development. Five states emerged as demonstrating comparable efforts: Arkansas, Colorado, New Hampshire, New York and Utah.

This Phase III report integrates findings from the prior phases with the high-level specifications and framework to meet the GMCB expectations for enhanced access and utilization of VHCURES through use of structured files. This report provides a “roadmap” of the design principles and high-level framework to increase user access to VHCURES as a valuable and instrumental analysis tool to quantify and analyze health care delivery in Vermont. This high-level framework is presented in five chapters, with associated discussion and recommendations, as follows:

- **Chapter 1** – General Principles and Considerations
- **Chapter 2** – High Level Framework
- **Chapter 3** – Methods for Meaningful Public Use Files
- **Chapter 4** – Recommendations to Address Different User Types
- **Chapter 5** – Recommendations Summary

The “actionable” summary list of recommendations incorporates current industry best practice for consideration to improve access and to meet the evolving needs of the GMCB and the health care stakeholder community for ongoing analytical power. Summary recommendations including priorities

for next steps are presented in Chapter 5, and address the following principles and foundational next steps:

High-Level Principles as Recommendations:

- Design data structures based on the operation of underlying process, rather than on the format of input documents.
- Maximize the amount of knowledge built into the data structures.
- Make individual records meaningful things to count.
- Separate clinical and financial information.
- Simplify the timing challenges inherent in the data.
- Improve the consistency of both individual and provider identifiers.
- Improve the utility of “support” files.
- Improve data quality / trust in data.
- Develop a mechanism to address the changes in what is included in VHCURES and potentially, simulate missing records.

Additional foundational recommendations include:

- Prioritize identity management approach proposed in the earlier VHCURES RFP.
- Integrate the Blueprint for Health’s primary care physician provider inventory.
- Incorporate key geographic and socio-economic data to support population analyses.
- Create a full validation report and process, making data available to payers prior to public use.
- Creation of an “initial public use” numerator file as detailed in Appendix 1 in compliance with safe harbor requirements to produce a simplified “one record per person” at regular intervals.
- Development of brief, online training support for first time and/or less experienced users for both VHCURES and VUHDDS data sets.
- Establish access “on ramps” for both files and for learning, in addition to flexible access to the enclave to reduce affordability barriers.

Through the process described herein, the GMCB will have a comprehensive reference inclusive of stakeholder input and current best practice designed to support the evolution and usefulness of the State of Vermont’s All-Payer Claims Database for ongoing delivery system evaluation and analysis.

I. Phase I and II Overview

The Green Mountain Care Board (GMCB) contracted with the Vermont Program for Quality in Health Care, Inc. (VPQHC) to expand access to and the usability of the Vermont Health Care Uniform Reporting and Evaluation System (VHCURES) and the Vermont Uniform Hospital Discharge Data System (VUHDDS) by producing a set of specifications for the development of data files tailored toward analytic interest. Phase I of this work included the review and

evaluation of previous reports and work products developed for the Green Mountain Care Board (GMCB). The purpose of this review was to:

- Summarize conclusions based on the review of files provided by the GMCB.
- Identify potential use cases and user populations based on existing and recommended work products.
- Identify outstanding informational needs (gaps) required to fully develop recommendations for data products, users, and/or use cases.
- Recommend the approach for closing identified gaps.

Phase II of this work consisted of stakeholder interviews to help to inform “informational gaps” and ascertain potential “use cases” for the analysis-ready data sets along with creation of an inventory of All-Payer Claims Databases (APCD) and Hospital Discharge Data Sets used in other states. Extensive analysis conducted throughout the Phase II process included highlights from key informant interviews and is summarized below:

Table 1 – Overview of Phase II Stakeholder Interview Findings

Use Case	Value	Gap
Looking at Total Cost of Care per Capita by Age, Gender, and Health Service Area (HSA)	Provides a rich understanding of how costs vary among areas of the state, after controlling for differences in age and gender mix.	Information on the uninsured and from small insurers and non-participating employers. Possible additional data sources: Medical Expenditure Panel Survey (MEPS). While MEPS does not contain a sufficiently large sample of Vermonters to make it directly usable, use of relativities such as Total Cost of Care (TCOC) for the uninsured compared to those with commercial insurance can be used to create estimates.
Comparison of Utilization between ACO-attributed and non-attributed populations and across time	Get an understanding of how the attributed and non-attributed populations differ. Determine whether there is a selection bias that is impacting effects on utilization and cost.	Need to replicate OneCare attribution.

Use Case	Value	Gap
Spending on Patients with Specified Chronic Illnesses Over Time	Inform targeted interventions that can reduce incidence and severity of chronic illness.	Need to create definitions of diseases that can be used in VHCURES. For example, patients with diabetes can often be more reliably identified using prescription data.
Integrating Information from Birth and Death Certificates with VHCURES	Much richer analyses of births by including both claims and information from the birth certificate such as Apgar scores and parental demographics.	Need to develop a way to match names of vital records with encrypted names in VHCURES.
Match VHCURES Inpatient Facility Claims with VUHDDS	Reliability testing of both data sources individually (ideally, perfect matching of insured Vermont resident care at Vermont hospitals). Further information on the “border crossing” phenomenon – who is in one data set and not the other.	Ability to match with no transparent personal identifiers in either data set. How reliable is matching on available info – zip, gender, date of birth (DOB), admission and discharge dates, possibly diagnosis.

The APCD inventory product identified five states that responded to the “best practice” survey VPQHC conducted demonstrating robust utilization of APCD to support state-based health care reform and improvement initiatives. These five states are Arkansas, Colorado, New Hampshire, New York and Utah.

II. Phase III Introduction and Overview

The intent of Phase III: Recommendation of Principles and File Domains, as delineated in the contract language, is to:

“... produce a report that includes a set of principles that, at a minimum, will include the framework and high-level specifications for how GMCB may produce a set of analytic files for a variety of use cases (e.g., research and analysis of social determinants of health, chronic conditions, public use files) that allow a broader audience of analysts to use the data.”

The Phase III report will present four chapters and related Appendices that will provide detailed discussion of the opportunities to broaden access and usability of the VHCURES database, under the following chapter headings:

- Chapter 1: General Principles and Considerations
- Chapter 2: High-Level Framework
- Chapter 3: Methods for Meaningful Public Use Files
- Chapter 4: Recommendations to Address Different User Types
- Chapter 5: Recommendations Summary

Chapter 1 – General Principles and Considerations

The Vermont Health Care Uniform Reporting and Evaluation System (VHCURES) is Vermont's All-Payer Claims Database (APCD). VHCURES and The Vermont Uniform Hospital Discharge Data System (VUHDDS) is Vermont's hospital discharge data set. Both are extremely valuable data resources for a range of efforts to understand, improve, and reform the state's healthcare system. While both data sets have provided valuable information for many years, to date their use has been limited to a small group of analysts with substantial technical experience and expertise. In an effort to expand the user base of each of these data sets, the Green Mountain Care Board (GMCB) seeks to design foundational data file structures that will support broader access for valuable partners who can utilize VHCURES and VUHDDS data more robustly. The problem statement summary for this undertaking is taken directly from the RFP document as follows: "In the database format, significant technical skills and resources are required to synthesize the data before they can be meaningfully analyzed. The GMCB wants to produce a set of data files designed for more immediate analysis by a larger number of people so that it may expand access to and the utility of its rich data resources. Broadening access to the data contained in VUHDDS, VHCURES, and other Vermont-level health care data sources will promote the quality of the data, enhance its value, and increase opportunities for meaningful research."

The VPQHC project team has conducted this project under a set of goals and principles intended to achieve the GMCB's goals. These project goals include:

- The design of a database should maximize the knowledge incorporated into that design to minimize the level of experience and expertise required of users.
- Data structures should replicate, to the extent possible, the system that they are designed to mirror.
- The best data design by itself is not sufficient to guarantee usability of the data – training and support are of equal importance.

VUHDDS and especially VHCURES in their current formats require substantial expertise to deal with such issues as multiple claims for a single clinical event, the multiple forms of payment for care, and the conflicting goals of completeness and timeliness of data. For example, a question such as "how much does a knee replacement cost?" requires knowledge of claims submitted by multiple entities (at a minimum, hospital and surgeon), the difference between charges, allowed charges, payments, and out-of-pocket costs, and the analytic consequences of claims often being paid well after the service has been provided.

Vermont has embarked on a reform of its healthcare system that will pay for value rather than volume. Basically, this reform separates the provision of a service from the payment for that individual service (fee-for-service model). A claims-based system is built on fee-for-service – each line of each claim has a payment calculated for it. In order to accurately reflect the reforms that have been undertaken, the data structures must separate clinical from financial information and reflect the multiple ways payments can be made (e.g. DRGs, episodes, and capitation).

Many potential users of VHCURES and VUHDDS have been frustrated by the complexity and size of the data sets. A range of support mechanisms are necessary to allow users to gradually gain expertise while their efforts are supported.

Chapter 2 – High Level Framework

This chapter describes the framework and high-level specifications on how GMCB could produce a set of analytic files for a variety of uses that allow a broader audience of analysts to use the data.

Introduction

This chapter discusses the goals and principles that will help to make GMCB data, particularly VHCURES, more accessible to analysts.

These goals include:

- Provide a roadmap for making VHCURES a more accessible and useful database.
- Provide recommendations for specific actions the GMCB can take to get results necessary to carry out its mission.
- Identify best industry practices.

These principles include:

- Design data structures based on the operation of underlying process, rather than on the format of input documents.
- Maximize the amount of knowledge built into the data structures.
- Make individual records meaningful things to count.
- Separate clinical and financial information.
- Simplify the timing challenges inherent in the data.
- Improve the consistency of both individual and provider identifiers.
- Improve the utility of “support” files.
- Improve data quality / trust in data.
- Develop a mechanism to address the changes in what is included in VHCURES and potentially, simulate missing records.
- Provide multiple means of access to data.

Design Data Structures Based on the Operation of Underlying Process, Rather Than on the Format of Input Documents

The most significant organizing concept in health care analysis is that of a clinical event – an interaction between a patient and one or more providers with a specified reason for that encounter (the diagnosis) for which one or more services are provided (procedures and revenue codes). Claims are nothing more than invoices submitted by each of the providers in order to receive payment. Implicit in a claims-based model are assumptions that are artifacts of the earliest health insurance models. These assumptions include:

1. Fee-for-service payment:
 - Payment will be made at the most detailed possible level.
 - Each service will have a charge, determined by the provider.
 - Each service will generate a payment, determined by existence of and terms of a contract between the provider and the payer and by the terms of the benefits specified in the contract between the payer and the beneficiary.
2. Under this payment model, a single clinical event may generate multiple claims
 - if multiple providers were involved (e.g. a hospital inpatient surgery),
 - if multiple payers were involved (coordination of benefits / other party liability), or
 - if the clinical event is of extended duration.
3. Useful information exists in the individual transactions between each payer and each provider.

These same assumptions are problematic in nearly all analyses of clinical care. The assumption that fee-for-service payment remains the dominant payment mechanism is becoming problematic as payment reform efforts progress with the implementation of many new initiatives. The second and third assumptions add enormous complexity to any effort to analyze care. One of the most straightforward ways to improve the utility of VHCURES would be to offer a version that summarizes a clinical event into a single record. While this action is not trivial, it would greatly simplify the usefulness of VHCURES, while offering the possibility of a smaller file/table.

Maximizing the Amount of Knowledge Built into the Data Structures

During the interviews conducted as part of Phase II of this project, several interview subjects identified barriers to broader use of VHCURES data. One of the most mentioned barriers was the complexity of VHCURES data, as discussed above. Some of the major factors that contribute to that complexity include:

- A “header-detail” data model (which replicates the structure of a standard claim).
- The possibility of claims from multiple providers for a single clinical event.
- The possibility of multiple claims from a single provider, processed by a single payer for an ongoing clinical event, such as a long hospital stay.
- The possibility of multiple claims for the same service, processed by different payers.
- The challenge of incomplete data for a specified service period (“run out”).
- Multiple coding systems for diagnoses and especially procedures.

- Complexity of payment systems, including both fee-for-service and alternative payment systems.

While there are techniques for addressing each of these as discussed in Appendix 1, the most important design consideration is to maximize the knowledge built into the data structure, while reducing as much as possible the level of expertise required of the user.

For example, each user should not have to develop their own way to count and summarize a clinical event (use the hospital claim? the surgeon claim? add the paid amounts? what if the primary diagnosis differs?).

Make Records Meaningful To Count

A main reason that VUHDDS is much easier to use for analyses than VHCURES is that each clinical event (in VUHDDS an inpatient stay or outpatient visit) is summarized in a single record in the main table¹. A question such as “How many knee replacements have been done in Vermont?” is straightforward to answer.

In VHCURES, answering that question would require deciding which records to count (hospital or surgeon) and filtering numerous records to find the single one to count.

Separate Financial and Clinical Information

In a claims database, there is usually financial information associated with each individual service line (room and board, operating room, medical supplies, etc.). Under the traditional fee-for service (FFS) payment, information for each line would include charges (set by the provider), “allowed charges” (based on a contract between the payer and the provider, if one exists), and deductible, coinsurance, and copayment (based on the patient’s specific benefits). Prior to the 1970s, this was the standard payment model for nearly all health care. In the late 1970s and early 1980s, the first effort to reform payments to hospitals was implemented. This was the Diagnosis-Related Group or DRG system. Under this model, payment was made at an aggregate, rather than individual service level. Payment was based on diagnoses, procedures, and sometimes on patient characteristics. Charges and volume of services (e.g., length of stay) were usually irrelevant. Medicare was an early adopter of this system, but many private insurers use it for some or all inpatient services. Because the payment amount is known prior to the payer processing the claim, this type of payment is called “prospective.”

A subsequent round of reforms applied this model to outpatient hospital care, but again only to the hospital’s portion of the payment.

It can be challenging to find these payments in VHCURES (which vary by submitting payer), but because both were designed to pay only the hospital share of care costs, the broad type of service is consistent and payment can be attached to a single claim, if not to individual lines.

¹ There is an additional table that provides revenue-code detail

The next round of reforms, “bundled payments” focused on expanding the payment to include both facility and professional services. For example, a surgeon and a hospital would share a single payment for a joint replacement. This payment model still applies to a single clinical event but there is no standard way to report the payment in a database whose structure is based on a traditional claim (multiple lines of service). Some payers report the payment on the first line of the claim, others report it on the last, and others create a special line just for reporting the payment amount.

Another payment reform addressed episodes of care. While this can be synonymous with a bundled payment in some cases, in others it expands the scope of payment to multiple dates of service as well as multiple providers. A good example of this is a payment for a joint replacement that also covers rehabilitation.

Each of the above payment models are connected in one way or another to provision of services. However, there are other alternate pathways for payment reform that completely disconnect payments from provision of services. The most well known example of this is capitation. Under a capitated model, a provider (or group of providers) receives a fixed payment for a defined period (most often monthly), for each patient who either chooses that provider (HMO / primary care) or who is assigned to that provider / group (ACO). Other variations of the non-service payments include continuing to pay some variant of fee for service (FFS) but settling the amount to a negotiated target and in its most extreme form, establishing a global budget (e.g., the Canadian hospital payment system).

In attempting to reflect the current mix of payment methods in a claims database, we must be able to manage payments that are attached to:

- individual services (FFS),
- a hospital stay (DRG),
- multiple providers (bundled payments),
- multiple providers and dates of service (episode payments), and,
- an individual for a specified time period (capitation).

It is clear that it would be very challenging to adapt the existing data structures to be able to reflect all the different payment models in a meaningful way. Perhaps even more significantly, it is the State’s intent to move away from any form of service-based payment and instead pay for value.

“The goals of payment reform in Vermont are to move away from volume-based payments toward a payment system that reinforces our efforts to improve the health of Vermonters, improve quality of care, and contain the rate of growth in health care costs. We are testing systems on a pilot basis with willing providers, across all payers -- including Medicaid and Medicare. We will evaluate the pilots to judge their applicability to broader populations of health professionals and patients².”

² <https://gmcboard.vermont.gov/payment-reform> (12/7/2020)

Simplify Timing Challenges Inherent in the Data

Final payment of a claim (“paid date”) does not often occur in the same month as the service is provided (“incurred date”). Typically, about 25 percent of dollars are paid in the same month, and 50 percent are paid in the next month. About 95 percent of dollars are paid within six months and 97 to 98 percent in 12 months³. This lag is the consequence of two different factors. The first is simply processing time. Providers take time to assemble and submit claims and payers take time to adjudicate them. The second factor is that claims may be adjusted after initial payment as a result of new information, such as the existence of other insurance coverage.

This pattern results in a conflict between completeness/accuracy of analysis and timeliness of analysis. It also may introduce a replicability issue when a query that specifies only incurred dates is run at different points in time.

Several different approaches can be taken to mitigate these problems. The first, used by Medicare, is to allow users access only to “final action” files. These are usually created by selecting claims with an incurred date during a calendar year and a paid date from the beginning of that year through six months of the following year. While this eliminates nearly all issues around completeness and consistency, it does not allow for timely analyses.

The second approach is the creation of what are called completion factors used in actuarial analysis at insurance companies. These are based on actuarial analyses of payment patterns (the creation of claims triangles) that establish historical distribution patterns of payment dates for a specified incurred month. This was the basis for the distribution referenced above. For example, for claims incurred in October of 2020, we would expect 75% of payments to have been made by the end of November, so we complete the reported payment amount by dividing by 0.75. While this approach, which is used for insurance rate filings, can provide timelier (estimated) data, it is complex and is not as useful for counting clinical events. For example, complex hospital claims take longer to process than simple office visits.

Neither of the above approaches directly addresses the issue of adjustments. These are particularly challenging in a multi-payer claims database because payers take different approaches to managing them. For example, Vermont Medicaid uses a “versioning” approach, where multiple versions of an adjusted claim exist in the raw data, and the data retrieval process is relied upon to select the correct version. One common approach among insurers to representing an adjustment is called “void and replace.” Under this approach, the original claim is retained with its original paid date and two new claims are created with a new paid date, one to reverse (negate) the original and one to reflect corrected information. The advantage of this approach is that any analysis that specified both a paid and an incurred period will produce consistent results regardless of how long after the period of interest it is run.

Note that because of the complexities of this issue, when the state’s current APCD vendor generates new claim files, it takes an “update in place” approach, which changes the original claim to reflect the new data.

³ Policy Integrity analysis of VHCURES data, April, 2018

Provider and Individual Identifiers

The ability to consistently identify patients and providers is essential in most types of research that are supported by VHCURES. Patient identifiers are necessary for any longitudinal analyses such as treatment paths and episode of care analyses. Provider identifiers are an important component of analyses such as price transparency and quality improvement.

Currently, individuals are identified in VHCURES by hashed⁴ name and social security number. The hashing is done by the payer that is supplying the data. While both of these fields are likely to be consistent over time within a payer's data (the main exception is a name change), both are problematic, especially when trying to assemble information across payers. For example, it can be difficult to track an individual whose source of coverage changes. The challenge with names is that a hash when the name includes a middle initial will be completely different from a hash of the same name other than the initial. Many organizations are moving away from using Social Security numbers as identifiers (including Medicare and Blue Cross) and many insured individuals are refusing to provide them.

These challenges are even greater when attempting to link VHCURES data with other sources. For example, linking commercial insurer records for a newborn with the newborn's birth certificate currently needs to rely on other attributes such as date of birth, gender, and ZIP code. While these matches can be fairly reliable, especially in ZIP codes that are sparsely populated, the risk of false matches cannot be ignored.

We recommend that the Board reconsider the identity management approach that was proposed in the VHCURES RFP several years ago. While evidence from other states suggests that the proposed approach is cumbersome and expensive, we believe that this more reliable way to track individuals would substantially increase the value of VHCURES.

The federal government has created a national mechanism to identify providers, called the National Plan and Provider Enumeration System (NPPES). Within that system, each provider, including both institutions and professionals, is assigned a National Provider Identifier (NPI). Any provider billing either Medicare or Medicaid is required to have an NPI. Some commercial insurers have adopted NPIs but others have not.

There are two limitations to NPPES. The first is maintenance. While providers are supposed to update their information when anything changes, this is not always done, especially with geographic information. The second is that the ability to accurately portray complex organizations is limited. For example, there is no simple way to identify all component parts of UVMHC (e.g., ambulance service, pharmacy) and all the physicians who are employed there.

We recommend consideration of a consolidated Vermont provider inventory. A good model for this is the registry of primary care physicians that the Blueprint maintains.

There are few identifiers in VUHDDS. Hospitals are identified by an assigned number. Physicians who provided care during a procedure or stay are identified by a number assigned by the hospital but as far as we know, there is no central index. Other than date of birth, gender, age, admission date and

⁴ Hashing is a form of encryption designed NOT to be decoded. In other forms of encryption, if someone knows the key, the original information can be recovered, while in hashing there is no key.

discharge date, there are no patient identifiers. This makes cross-hospital tracking such as transfers difficult.

Support Files

While much of the discussion around VHCURES refers to it as a claims database, it is composed of four broad types of files:

1. claims
2. population/membership/eligibility
3. pharmacy
4. provider

In addition to these, there are several “support” files, including: a. diagnoses, b. procedures, and c. geographic, with the possibility of a new type – d. linkage. Each of the existing files presents an opportunity to make VHCURES easier to use and more valuable.

For many analyses, the population file is as valuable as the claims file. In other words, in the calculation of rates, the denominator is as important as the numerator. The current population file shares some of the problems that make the claims file a challenge to use. In particular, rather than having one record per person for a specified time period (usually one month), these files have one record per person per time period for each payer. For example, an individual with both Medicare and Medicaid would appear twice for January of 2020. Further, while encrypted identifiers on the claim and person records for a specific payer are reliably the same, there is no reliable way to link person (or claim) records across payers. Even more problematically, there is no reliable way to link the records of an individual who switches payers, making longitudinal analyses challenging.

The pharmacy files contain information on every prescription that is filled and that is paid for by a third party⁵. While pharmacy data can be extremely valuable (e.g., for identifying individuals with insulin-controlled diabetes), there are major obstacles to integrating medical and pharmacy data. The first of these is that the pharmacy data cannot be directly linked back to any medical claims. One of the possible reasons to connect these two files is to explore the reasons for which a drug is prescribed (NOT a part of the pharmacy record). While it can be assumed that a prescription filled within a few days after a visit to a prescribing professional was in response to that visit, refills cannot be connected that way. In addition, there are a number of coding schemes to drugs, such as NDC⁶ and the “Orange Book”⁷.

Provider files have long suffered from the lack of a reliable cross-payer way to identify providers. Typically, we are interested in both the rendering provider (care actually provided) and the billing provider (to whom the check was written). For many years, each payer had its own idiosyncratic way of identifying providers. In 2006, the Centers for Medicare and Medicaid Services (CMS) began issuing nationally standard provider identification numbers in the hopes of simplifying this problem. NPIs were a major step forward, but are limited in a couple of ways. First, it is up to the provider to update the record in the NPI database. This does not always happen. Second, the NPI system is not well suited to

⁵ Subject to the same limitations as the rest of VHCURES, e.g., no federal employees

⁶ <https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory>

⁷ <https://www.fda.gov/drugs/information-health-care-professionals-drugs/electronic-orange-book>

complex organizations. For example, it is difficult to aggregate data for the UVM Health Network. Third, NPI data do not reflect the activities of providers who work for multiple organizations.

One valuable but labor-intensive alternative is the Blueprint for Health's provider registry. This data source provides timely and accurate information on a subset of providers working in Vermont. Consideration should be given to integrating the registry into VHCURES.

Diagnosis coding is a central resource in nearly all health care analyses. Far and away the most prevalent diagnosis classification system is the International Classification of Diseases (ICD). While it is an essential part of most analyses, it does have several limitations. The most significant is the minimal coverage of mental disorders. A distinct classification scheme, the Diagnostic and Statistical Manual of Mental Disorders⁸ (DSM) is maintained by the American Psychiatric Association. Fortunately, efforts are underway to bring the two systems into greater coordination.

There are other diagnostic systems that have been created to address the limitations of ICD, but they are not in broad use. The most significant of these is SNOMED, which includes both diagnostic and procedural codes.

The dominant procedural coding system in the US is the Current Procedural Terminology (CPT) system. CPT is owned by the American Medical Association and requires a license to use. It is extremely comprehensive. However, because it is proprietary, there is no dictionary available to users of VHCURES other than state agencies.

Related Records to Place

Geography is becoming a much more important component of health care analysis. Geographic analysis can provide basic information such as patterns of travel for care and physical access to services. With increasing interest in determinants of health, it is also possible to use the characteristics of a geographic area (educational level, median income, etc.) as an estimation of the relationship of place to health.

Currently, geographic information in both VHCURES and VHDDS is limited to ZIP code (for all payers) and Town (all payers except Medicare). A question has been raised about whether the absence of Town in the Medicare files has any impact on their ability to support geographic analyses. Because both fields are typically derived from mailing addresses, it is highly likely that ZIP and Town are redundant.

However, there is a larger question to consider. Often neither the ZIP or Town fields reflect the actual town of residence. For example, the town of Berlin, where Central Vermont Medical Center is located, receives mail through either the Montpelier (05602) or Barre (05641) Post Office. There is no ZIP that is specific to Berlin.

This problem was first identified by Vermont Cooperative Health Information (CHICV), the organization⁹ that first used Vermont hospital discharge data for utilization analyses. The crux of the problem was at that point in time, the most detailed reliable population data was at the physical town level, while

⁸ <https://www.psychiatry.org/psychiatrists/practice/dsm>

⁹ The Vermont Cooperative Health Information (CHICV). CHICV operated in the late 1970s and early 1980s, and was responsible for the original hospital monograph series.

location information was based on ZIP. CHICV created what they called ZIP Towns, which were groupings of ZIP and Town designed to minimize the impact of potentially inaccurate population data.

While it would be straightforward (but tedious) to update ZIP Towns, the geographic utility of both VUHDDS and VHCURES would be improved if location could be more, rather than less specific.

There is a wealth of socio-economic data to support these kinds of analyses. All attempts to make access to this data in a new analytical framework would be a great addition.

The Vermont Center for Geographic Information is the official state repository for boundary files. The following are recommendations for these data sets. In addition, very inexpensive up-to-date and easy to use socio-economic data is available from a number of vendors. The following are the key geographic and socio-economic databases that warrant a place in the usability of VHCURES in any analytical framework. The following are the key geographic databases to support population analyses. There are a number of geography-based boundary files listed below.

HSA-Health Service Areas: Health Service Areas are an aggregation of Vermont towns that the Blueprint for Health has created to manage and report on primary care in the Community Centers. These differ from VDH's HSAs because of differing Vermont towns in each. There is a need for a "look-up" file to crosswalk the towns that belong to either Health Service Areas, Hospital Service Area, and VDH Health Districts.

HSA- Vermont Hospital Service Areas: Hospital service areas (HSAs) are organized by towns and are based on inpatient discharges where the diagnosis indicated the need for immediate care. Plurality inclusion rules result in towns being assigned to the HSA corresponding to the plurality of discharges. HSA analyses are used to compare data for residents of 13 geographic regions of Vermont who were provided inpatient and select outpatient services in any Vermont, New Hampshire, New York, or Massachusetts hospital. HSAs are defined by the Vermont Department of Health (VDH). Information on HSAs can be obtained online at the [Vermont Open Geodata Portal](#). The VDH Public Health Statistics program periodically updates HSA GIS data and was last updated in 2016. The VDH maintains this GIS data for public health planning and emergency response purposes and reports data on this basis.

Vermont Health Districts: The Office of Local Health carries out the Vermont Department of Health's mission to protect and optimize the health and wellbeing of its citizenry by supporting a statewide public health delivery system. This system provides essential public health services and promotes building local infrastructure to address public health needs of Vermonters from the community perspective. Access this information at the Vermont Open Geodata portal for [Vermont Health Districts](#).

County and Town Boundaries: The Vermont Center for Geographic Information's (VCGI) BNDHASH (data set name) Boundary data layer contains all village, town, county, RPC, and state boundaries into a single data layer. VCGI has assured vertical integration of these boundaries and simplified maintenance. BNDHASH also includes annotation text for [town, county, and RPC names](#).

ZIP 5: A ZIP Code Tabulation Area (ZCTA) is a statistical geographic entity that approximates the delivery area for a U.S. Postal Service five-digit ZIP Code. ZCTAs are aggregations of census blocks that have the same predominant ZIP Code associated with the addresses in the U.S. Census Bureau's Master Address File (MAF). ZCTAs do not precisely depict ZIP Code delivery areas, and do not include all ZIP Codes used for mail delivery. Defining the extent of an area is necessary to accurately tabulate census

data for that area. ZCTAs are generalized area representations of U.S. Postal Service (USPS) ZIP Code service areas. Simply put, each one is built by aggregating the Census 2000 blocks, whose addresses use a given ZIP Code, into a ZCTA, which gets that ZIP Code assigned as its ZCTA code. They represent the majority USPS five-digit ZIP Code found in a given area. For those areas where it is difficult to determine the prevailing five-digit ZIP Code, the higher-level three-digit ZIP Code is used for the ZCTA code. TIGER, TIGER/Line, and Census TIGER are registered trademarks of the U.S. Census Bureau. ZCTA is a trademark of the U.S. Census Bureau, and can be accessed [VT Zip Code Areas](#) of the Vermont Open Geodata Portal.

Socio-Economic Datasets (essential for Population Health Analyses): There are multiple sources of socio-economic data that can be combined with VHCURES analytical data sets. To ensure consistency of analysis we believe that it would be valuable to develop standard for Vermont's socio-economic data used for any Vermont analyses to ensure consistency. These data sets typically originate at the national and state level from the Census, [surveys and spending data](#).

One of the more significant areas for development is that of linkage files. These files could maintain each of the identifiers associated with an individual over time. In addition to potentially solving the problems described above (inconsistent identifiers across payer and time within VHCURES), linkage files could open access to data systems external to VHCURES, including VUHDDS and vital records. The information from these files could add enormously to the utility of VHCURES. For example, birth certificates include information on parental education and a summary of newborn health, called the APGAR score. In addition, linking VUHDDS and VHCURES could provide a way to validate shared data elements.

For example, using Esri's Community Analyst data and analysis subscription enables the user to get information for over 2,000 data elements for the exact area needed—including standard geographies (down to the Census block group level), hand-drawn shapes, or rings or drive times around a location. It includes access to thousands of demographic, Census, health, crime, and business variables to formulate better policy decisions. It goes beyond basic demographics to allow an understanding of the behaviors and preferences of people living in any area using their detailed segmentation profiles.

In the GCMCB's Draft *Health Resource Allocation Plan* conducted in 2019, a pilot End-Stage Renal Disease (ESRD) geographic-based analysis was conducted using ESRD data from VHCURES. The pilot example illustrated some of the issues related to such an analysis. The first was accessing enough patient data by ZIP code so that confidentiality requirements could be met. This required looking back over a 3-year period. The findings were of interest for identifying the frequency and travel distance a patient undertook for treatment. It provided useful information in formulating locations for locating more dialysis beds based on population location information.

The VDH has a [Vermont Social Vulnerability Index \(SVI\) interactive map applications](#) based upon American Community Survey data. The application contains a number of vulnerability measures above the 90th percentile for Vermont census tracts. For each of the vulnerability measures, census tracts in the 90th percentile of vulnerability were assigned a flag. This SVI exhibits the sum of all flags for each census tract. There are a total of 16 measures in the SVI "social well-being" dataset that can easily be used for comparative analysis with VHCURES tables.

Improve Data Quality / Trust in Data

The usefulness of any data system is strongly influenced by user perceptions of the accuracy and reliability of the content of that system. One of the interesting questions in a derived database such as VHCURES is what should be compared to establish accuracy. Ideally, medical information in VHCURES (diagnoses and procedures) should agree with the provider's medical records, while those fields might be subject to change during the provider's claim submission process, the payer's adjudication process, and the data warehouse ETL (extract, transform, and load) processes.

The lack of access to medical records rules out validation from the source (at least until a unification of VHCURES with electronic medical records). At this time, the best we can do is verify the end-user version of the database against information provided by the payers. Some preliminary efforts to do this have been made, with results that raised some concerns.

Both Medicaid and Blue Cross have raised concerns about differences between their records and reports based on VHCURES data. While there may be legitimate reasons for these differences, it is essential that the major suppliers of data to VHCURES understand those differences and, as a result of that understanding, accept results based on VHCURES.

This will require work in two areas. The first is **enhancing the intake and edit process** (submission of raw data from payers). As we understand the current system, while there are numerous edits done on the data, the only feedback that payers receive is when an element fails an edit. We recommend the creation of a full validation report, developed in consultation with the payers, and made available to the payers prior to making the data available to users.

The second is **developing a method of making their own data freely available to payers**. Allowing payers to run queries on their data as it exists in VHCURES may help to eliminate their concerns about data quality. Of course, it is essential that however access is granted, payers must not be able to see any data other than their own.

Weighting

Two significant challenges in using VHCURES to study health care received by Vermonters are both related to who is included in the data and who is not. From the time VHCURES was first created, small insurers have not been required to submit data. Federal employees have also been excluded (based on federal law). While this is a small number of lives (estimated at 10 to 20 thousand), it has always led to underestimation such as total spending on behalf of Vermonters.

Of much greater consequence is the gradual erosion of participation by private payers in VHCURES. The most significant influence on participation was the March 2016 US Supreme Court decision in *Gobeille vs. Liberty Mutual*, which invalidated Vermont's mechanism for collecting information from self-insured ERISA plans. Beyond this, there appears to have been a slow but steady decline in the number of insured lives in VHCURES. Appendix 2 is an analysis done by Policy Integrity in 2017. Note that the observed decline has stabilized since the *Gobeille* decision.

Another limitation is that utilization of health care by the uninsured and by insured individuals who choose to pay for some services entirely out of pocket (self-pay) is not included in VHCURES (for obvious reasons).

To provide the most accurate possible picture of health care utilization and spending by Vermonters, a method must be developed to account for both the population that was never in VHCURES and the population that has been lost over time. One possible (but not simple) way of doing this would be to “weight” records. Weighting data is a statistical way of saying that each individual record may contribute a multiple of its content. For example, suppose we want to increase the total spending by current beneficiaries of self-insured employers to adjust for a 10% decline in the number of self-insured lives in the database. Over-simplifying things a little, we could weight each record by 1.111 (100% / 90%). A claim that had a payment of \$100 would, after weighting, contribute \$111. While this makes the database more complicated to use, it improves accuracy.

It is also possible (but may be even more difficult) to create synthetic claims on behalf of the uninsured. This would require adjusting data from a national survey (MEPS – the [Medical Expenditure Panel Survey](#)¹⁰) to Vermont conditions and creating records based on that adjustment.

Other Data

We suggest that as part of any effort to make VHCURES useful to a wider range of analysts, the GMCB consider inclusion of some additional data. Ideally, identification of these new elements would be done in consultation with existing and potential users, but three types of data should certainly be included in consideration.

The first is **demographic data** – specifically elements like education, income, and race. While it may not be possible to collect these elements at a true individual level, it is certainly possible to include them in any geography-based analysis. For example, there is a sizeable difference in health care spending per capita between the Rutland and Chittenden areas of Vermont. Historically, that difference has been attributed primarily to differences in how medicine is practiced (age differences also contribute). Focusing on changing practice patterns may not achieve the desired changes if a significant portion of the difference in health spending can be explained by differences in demographics.

The other two areas for which there has been some demand in the past are **insurance premiums and benefits**. While payers have been reluctant to provide this information in the past, it can provide valuable information on how much households actually spend on healthcare rather than how much is spent on their behalf by insurers.

Multiple File Formats

There are two main approaches to maintaining an analytic file of the magnitude of VHCURES. The first is in a relational database, e.g., Oracle or Amazon’s Redshift. The second is in a statistical package such as

¹⁰ <https://www.meps.ahrq.gov/mepsweb/>

SAS, STATA or R. Each has advantages and disadvantages. The Board’s “analytic enclave” is built using a relational database. We believe that a number of current users are using statistical software packages.

The two key activities in health services research are data management and data analysis. The two recommended file structures reflect the different strengths required for each of these.

Relational databases excel at the management of large volumes of data and are particularly effective in the areas of sub-setting (find all the claim records for 20- to 30-year-old women who had hysterectomies in 2019) and consolidating information across tables (count primary care visits in the claim file by the location of the provider’s office, in the provider file – a common Blueprint analysis). However, most relational databases do not have extensive statistical tools. For example, it would usually be necessary to extract the data and use another tool if we wanted to explore the statistical relationship between age and health care spending.

Statistical packages such as SAS or STATA offer a rich set of analytic tools, but are not as efficient in data management. One form of analysis that can be done easily in a statistical package but is extremely difficult in a relational database is time between services. This is because record by record processing is a challenge in databases. An example of this type of analysis is identifying episodes of medically-treated low back pain. A common definition of an episode is all care received until a six-month period passes with no back-related care (the wash-out period).

A full-featured analytic environment would offer both a database and a statistical package, ideally with some level of integration between the two. For example, a query embedded in code for a statistical package could be passed to the database and the response could then be processed statistically.

Prior to the creation of the enclave, any user who wished to make use of VHCURES was given flat files. A flat file is a simple text file, with data elements either arranged in fixed-width columns or separated by a specific character. Each user would have to create scripts to load the data into the software of preference.

The enclave is a major improvement over versions of the data managed by multiple users. However, as far as we know, the enclave relies purely on a database system. It would be beneficial to provide better support to users who prefer a statistical package, either making those packages available in the enclave or making “transport” files available to users. Transport files are files created in a format that can be recognized by most of the major statistical software packages.

Chapter 3 - Methods for Meaningful Public Use Files

This Chapter describes proposed methods to aggregate data into meaningful and compliant public use files, recommendations on the approach to linkage across files, and guidance about basic structure.

Introduction

Public use files are designed to balance ease of use and utility with privacy and confidentiality concerns. The Centers for Medicare and Medicaid Services (CMS) has developed a public use files system that enable researchers and policy makers to evaluate geographic variation in the utilization and quality of health care services for the Medicare fee-for-service population. Their Geographic Variation Public Use

File includes demographic, spending, utilization, and quality indicators at the state, hospital referral region (HRR) level, and county level. To support this effort, CMS has created a three-tier model of data file types¹¹. These are:

1. **Identifiable Data Files (IDFs)** — IDFs contain Personal Health Information (PHI) and/or Personal Identification Information (PII) and are only available to certain stakeholders. IDFs are available as custom extracts that can be shipped to a requester or accessed virtually. Requests for IDFs generally require a Data Use Agreement (DUA) with CMS.
2. **Limited Data Set (LDS)** — LDS files also contain PHI, but they do not contain specific direct identifiers as defined in the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. LDS files are available for research use. All requesters must complete a DUA with CMS.
3. **Public Use Files (PUFs)** — PUFs (also called non-identifiable data files) do not contain information that could be used to identify individuals. In general, PUFs contain aggregate level information on Medicare beneficiary or provider utilization. PUF requests do not require a DUA.

Eliminating nearly any possibility of identifying an individual is an essential requirement in the creation of public use files. HIPAA identifies two ways of establishing that a file has an ***acceptable level of identification risk*** or - ***expert determination*** and ***safe harbor***¹². Creating a HIPAA-compliant public use file from VHCURES¹³ that complies with the safe harbor model requires the blurring or elimination of several variables, including:

1. Names (currently encrypted in VHCURES)
2. Geographic information
 - a. No identifiers below state level with one exception – three-digit zip codes are permitted if the area encompassed by that three-digit code has a population of 20,000 or more according to Census Bureau data. Note that the only three-digit code that does not include 20,000 people in Vermont is 059.
3. All dates, including date of birth, admission and discharge dates, dates of service, and all ages over 89. Ages over 89 may be collapsed into a code for “90 and over.”
4. Social security numbers (currently encrypted in VHCURES)
5. Health plan beneficiary identifiers
6. Account numbers

Because of the limited utility of a public use file created from VHCURES that is fully safe-harbor compliant, we recommend that the expert determination approach also be considered. While the expert determination process is substantially more complicated than the safe harbor approach, the approach we recommend below relies on it.

¹¹ <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/Overview>

¹² <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>

¹³ This list does not include elements that are not currently included in VHCURES

While not explicitly discussed in the referenced CMS documents, similar considerations apply to tabulations derived from VHCURES data. A simple rule of thumb in tabulations is to suppress any cell with a count less than a specified value. However, if a small number of cells in a table are suppressed, actual values may be re-created using row and column totals.

While re-identification is the most commonly identified issue in the creation of public use files, usability for more naïve users can be equally challenging.

Proposal – First Step

We recommend creating an “initial public use numerator file” based on the “one record per clinical event” file discussed in Appendix 1. This file should be modified to comply with safe harbor requirements – particularly age recoding and three-digit ZIPs. In addition, the Board should consider reducing the level of detail contained in the diagnosis and procedural fields. For example, actual diagnoses could be replaced by their chapters in the ICD-10 system. For example, “Hepatitis A Without Hepatic Coma” (B15.9) could be replaced by its chapter, “Certain Infectious and Parasitic Diseases.”

These diagnoses should be paired with a simplified “one record per person (per month or possibly per year) that includes simplified primary source of coverage.

It is likely that the proposed public use files will be one of the best introductions to health care claims data in Vermont. In support of this, we recommend development of a brief online training video(s) designed for first-time users of both the VHCURES and VUHDDS public use files.

Subsequent Possibilities

Assuming there is sufficient GMCB interest in public use files the next step is to explore providing additional detail. Here, the key challenge will be in preventing the additional detail from increasing the probability of re-identification above acceptable levels. We recommend one possible way to minimize this risk would be to create a series of files, each one of which has *enhanced detail* in one field. For example, one version might have *full diagnosis information* while another might have *enhanced geographic information*.

If this approach is followed on full datasets, there is certainly a risk that a user could request multiple files and reassemble the details. To prevent this, we recommend creating the detail files as independent (non-overlapping) random samples. Creating ten percent (10%) samples would allow the creation of ten (10) distinct enhanced public use files. Implementing this approach would require expert determination.

Chapter 4 - Recommendations to Address Different User Types

Introduction

There is no one-size-fits-all approach that will meet the needs of all potential data users. There is a wide range of expertise, interests, and a database or statistical file that can be used by a relatively naïve user

will differ from one that will support an experienced analyst. Additionally, users will often differ in their analytic / research goals, their data management capabilities, and the structures of the organizations for which they work.

Recommendations

We recommend establishing two “on-ramps” for VHCURES and VUHDDS users. These would allow users to learn and use these files gradually. There would be two distinct on-ramps – one for files and one for learning. In addition to these on-ramps, we recommend options for data management, including both physical copies of data files and more flexible access to the enclave.

The file on-ramp would allow the user to progress from summary data to full access. The initial entry point would be downloadable summary files similar to the *Patient Origin and Enrollment Trend* reports currently available on the GMCB website, but with expanded documentation covering source (including code), possible uses, and caveats. Documentation for each file should be in a standardized format.

The next step could be graphic tools, such as Tableau or Microsoft BI. The GMCB’s Tableau version of the Patient Origin report is a good example.

Step three could be parameter-driven scripts. The [Census Bureau](#) has been particularly innovative in creating front ends for their data. For an example, see figure 1 below:

Figure 1: Census Bureau Data Usage Screen

The screenshot displays the Census Bureau's data portal. The browser address bar shows the URL: https://data.census.gov/cedsci/table?q=United%20States&g=0100000US_040000US0&tid=ACSDP5Y2018.DP05&hidePreview=true. The page title is "ACS DEMOGRAPHIC AND HOUSING ESTIMATES". The survey/program is "American Community Survey", the table ID is "DP05", and the product is "2018: ACS 5-Year Estimates Data Profiles".

The interface includes a search bar with "United States" entered. Below the search bar, there are tabs for "2 Geos", "Years", "Topics", "Surveys", and "Codes". The "2 Geos" tab is selected. The table shows data for the United States and Vermont.

Label	United States	Estimate	Margin of Error	Percent	Percent Margin of Error	Vermont	Estimate
SEX AND AGE							
Total population		322,903,030	*****	322,903,030	(X)	624,977	
Male		158,984,190	±6,691	49.2%	±0.1	307,973	
Female		163,918,840	±6,689	50.8%	±0.1	317,004	
Sex ratio (males per 100 females)		97.0	±0.1	(X)	(X)	97.2	
Under 5 years		19,836,850	±4,121	6.1%	±0.1	29,676	
5 to 9 years		20,311,494	±21,979	6.3%	±0.1	32,524	
10 to 14 years		20,817,419	±23,183	6.4%	±0.1	33,950	
15 to 19 years		21,204,226	±7,897	6.6%	±0.1	43,222	
20 to 24 years		22,286,970	±7,276	6.9%	±0.1	45,910	
25 to 34 years		44,567,976	±6,698	13.8%	±0.1	72,814	
35 to 44 years		40,763,210	±7,531	12.6%	±0.1	70,550	
45 to 54 years		42,589,573	±8,371	13.2%	±0.1	86,301	
55 to 59 years		21,611,374	±21,797	6.7%	±0.1	49,272	
60 to 64 years		19,675,357	±20,514	6.1%	±0.1	47,208	
65 to 74 years		28,535,419	±5,200	8.8%	±0.1	68,256	
		14,497,679	±13,915	4.5%	±0.1	31,606	

As user skills continue to develop, the Board could have simple summarized files or data base tables created. Rather than one record per instance of care, a record could include category variables (e.g., age, gender, HSA, and diagnosis chapter) and a count variable. This would be similar to how CMS makes its data available.

In a similar way, training should allow users to progress in parallel to the file ramp. In addition to staged training, several of our interviewees suggested ways to share information, including both specific code and definitions (“how do I identify a person with diabetes?”). Brief on-line training videos were also mentioned.

As mentioned above, management of these files, particularly VHCURES, is not trivial. In order to physically host VHCURES, a powerful file server (as opposed to a PC) is essential. Total file size for all components of VHCURES is well in excess of 100 gigabytes. Substantial experience in large file management (and data security) is essential. We believe the “Enclave” model that currently supports the Board, is an excellent alternative. Data are managed by an experienced organization and individual users purchase “seats.” NORC, as an early pioneer to the enclave concept, provides an [introduction](#) for users.

The main limitation to the Enclave that currently supports the Board is that seats must be purchased on a monthly basis. For users with occasional analytic needs, this is a major barrier.

To get around this barrier, we recommend exploring two possible options. The first would be a shorter purchase period. This could make a seat more affordable.

The second is to create a list of approved users available to support analyses. These approved users would be required to go through a similar approval process to what is necessary for obtaining a DUA, focusing on data management and data security. The users would maintain a seat and conduct analyses for clients. In order to address HIPAA and other issues, all products would be submitted to the Board for approval prior to being seen by clients. The Board might also choose to review requests before authorizing the ability to proceed.

Chapter 5 – Recommendations Summary

As described initially, this report provides a “roadmap” of design principles and a high-level framework that will improve both access to and the utility of VHCURES for ongoing analysis. The principles discussed in the previous chapters functionally serve as recommendations, and are summarized here with a brief explanation of considerations for redesign planning.

Design data structures based on the operation of underlying process, rather than on the format of input documents. The design of the current data structures, especially in VHCURES, reflect the structure of claims, rather than the underlying clinical activities. For example, there are multiple claims for a single inpatient surgery. One of the most straightforward ways to improve the utility of VHCURES would be to offer a version that summarizes a clinical event into a single record. While this action is not trivial, it would greatly simplify the usefulness of VHCURES, while offering the possibility of a smaller file/table.

Maximize the amount of knowledge built into the data structures. Users should not have to know how to consolidate multiple claims or deal with partial information for the most recent period of time. One of the most mentioned barriers was the complexity of VHCURES data related to the lack of a summary of clinical events into a single file. The bulleted items on page 8 details specific steps to maximize the amount of knowledge built into data files. Data files should be structured and ready for straightforward, meaningful analyses with clearly defined parameters that trained users can access without difficulty.

Make individual records meaningful things to count. Neither claims nor claim lines in themselves are meaningful for analysis. A question such as “How many knee replacements have been done in Vermont?” is straightforward to answer. In VHCURES, answering that question would require deciding which records to count (hospital or surgeon) and filtering numerous records to find the single one to count.

Separate clinical and financial information. As fee-for-service payment models become less and less prevalent, the utility and accuracy of structurally connecting individual claim lines with payment becomes more and more problematic. How much was paid for a service when the provider is paid under a capitated arrangement? In attempting to reflect the current mix of payment methods in a claims database, we must be able to manage payments that are attached to:

- individual services (FFS),

- a hospital stay (DRG),
- multiple providers (bundled payments),
- multiple providers and dates of service (episode payments), and,
- an individual for a specified time period (capitation).

Simplify the timing challenges inherent in the data. There are two distinct timing challenges with claims data. The first is that less than half of claims are paid in the same month as the service was provided in. This means that while a data model that is based on paid date (as is VHCURES) will be timely for paid analyses but not for service date analyses. The second is that claims are sometimes “adjusted” well after the initial payment date.

Improve the consistency of both individual and provider identifiers. While information on individuals is usually consistent within a payer’s data, it is challenging to connect individuals across payers. Nearly all providers have a nationally- consistent identification number, required to bill Medicare and Medicaid. However, the information associated with the provider (e.g. practice location) is not always kept up to date. In addition, there is not always a simple way to connect individuals with organizations, especially complex organizations. Patient identifiers are necessary for any longitudinal analyses such as treatment paths and episode of care analyses. Provider identifiers are an important component of analyses such as price transparency and quality improvement.

Improve the utility of “support” files. Many different coding systems are used in healthcare. Not all of these are documented within either VHCURES or VUHDDS. Improvement of these “support” files should be considered, including: diagnoses, procedures, and geographic, with the possibility of a new type – linkage. Improvement of each of these files presents an opportunity to make VHCURES easier to use and more valuable. Consideration should also be given to integrating the Blueprint for Health’s provider registry to provide accurate information on the subset of providers working in Vermont.

Improve data quality / trust in data. Concerns about data quality have been raised by two of the largest providers of data to VHCURES. An investment must be made to provide a more robust data editing process and submitters must have the opportunity to audit their own data.

Develop a mechanism to address the changes in what is included in VHCURES and potentially, simulate missing records. Due to several factors, but primarily to the 2016 U.S. Supreme Court case *Gobeille v. Liberty Mutual Insurance Company*, the total number of lives included in VHCURES has declined substantially. Further, there are small populations that are excluded, most notably federal employees and clients of small insurance companies. Unless accounted for accurately, these sources of undercount can bias analytic results.

Provide multiple means of access to data. Currently, state agency access to VHCURES is through a “data enclave,” provided by the state’s data management vendor. This approach is an excellent one, providing high levels of performance and data security. One large external user manages its own instance of the VHCURES database. Potential external users who do not have the capacity or expertise to manage a large database currently have no way to access VHCURES. Ideally, a mechanism should be identified under which infrequent users can access the data through the enclave.

In addition to the summary statements above, the following recommendations are identified as foundational steps to consider to improve access to and the overall utility of the VHCURES data set:

- Consider the identify management approach proposed in the VHCURES RFP from 2015. (Page 11)
- Integration of a consolidated Vermont provider inventory, such as the registry of primary care physicians maintained by the Blueprint for Health. (page 12)
- Incorporate key geographic and socio-economic data that is inexpensive, up-to-date and easy to use. Some key geographic databases to support population analyses. A number of key geography-based boundary files are listed on pages 14-15.
- Creation of a full validation report, developed in consultation with payers, including a process to make data available to payers prior to making data available to users. (Page 16)
- Creation of an “initial public use numerator file” as detailed in Appendix 1 to comply with safe harbor requirements. Consideration should be given to reducing detail in diagnosis and procedure fields with ICD-10 chapter level information. These diagnoses can then be paired with a simplified “one record per person (per month or possibly per year) that includes simplified primary source of coverage.
- Development of brief, online training videos designed to support first-time users of both the VHCURES and VUHDDS public use files (page 20).
- Establish two “on-ramps” for VHCURES and VUHDDS users, which would allow users to learn and use these files gradually. There would be two distinct on-ramps – one for files and one for learning. In addition to these on-ramps, we recommend options for data management, including both physical copies of data files and more flexible access to the enclave (page 21). To get around this barrier, we recommend exploring two possible options. The first would be a shorter purchase period. This could make a seat more affordable (Enclave, page 22).

Appendix 1 – Summarization of Claims into Clinical Events

In this discussion, a “clinical event” is one that includes services by one or more providers during a single day or a continuous period. Clinical events include a visit to a physician’s office, an outpatient surgery, or an inpatient surgery. For initial work on VHCURES, this distinguishes events from episodes. An episode can be defined as the set of clinical events that treat a single diagnosis over time. Examples of episodes include ALL care associated with a knee replacement, including subsequent physical therapy or a sequence of physical therapy visits for low back pain. We believe that ultimately the database should support both event and episode care, but event analysis is a simpler first step.

When creating a clinical event, several questions must be answered:

- What is the diagnosis for which care is being provided?
- Is the event on a single date or a continuous range of dates?
- How many providers were involved?
- How many payers were involved?
- What was the care setting?

Some examples are provided below:

Professional office visit

This is the simplest event to create. Aggregation will only involve multiple payers (necessary for accurately calculating patient liability)

Outpatient surgery

In this case, there will likely be both institutional (hospital or outpatient surgical facility) and professional (surgeon, possibly anesthesiologist) claims. Patient ID and date of service will be used to aggregate.

Inpatient surgery

Similar to outpatient, but over a range of dates. The hospital claim is the anchor with any professional claims during the data range of the hospital stay being included.

Appendix 2 – Enrollment Report

Analysis of Commercial Enrollment Data in VHCURES

10/6/2017

Introduction and Summary

The accuracy of information in VHCURES is a critical factor in the production of reliable analyses such as trends in per-member per-month (PMPM) spending for use in the evaluation of the all-payer model. This analysis provides a look at the quality of commercial enrollment data for medical coverage.

The analysis reveals a steady long-term decline in the number of commercially insured individuals (including self-insured) included in the database. From a peak in early 2008, total commercial enrollment, excluding behavioral health plans¹⁴, has dropped by a little under 50 percent. Looking at BCBSVT and the remainder of the commercial market, BCBSVT enrollment has risen about 63 percent from 1/2008 to 6/2017, while all other commercial enrollment has declined by over 88 percent.

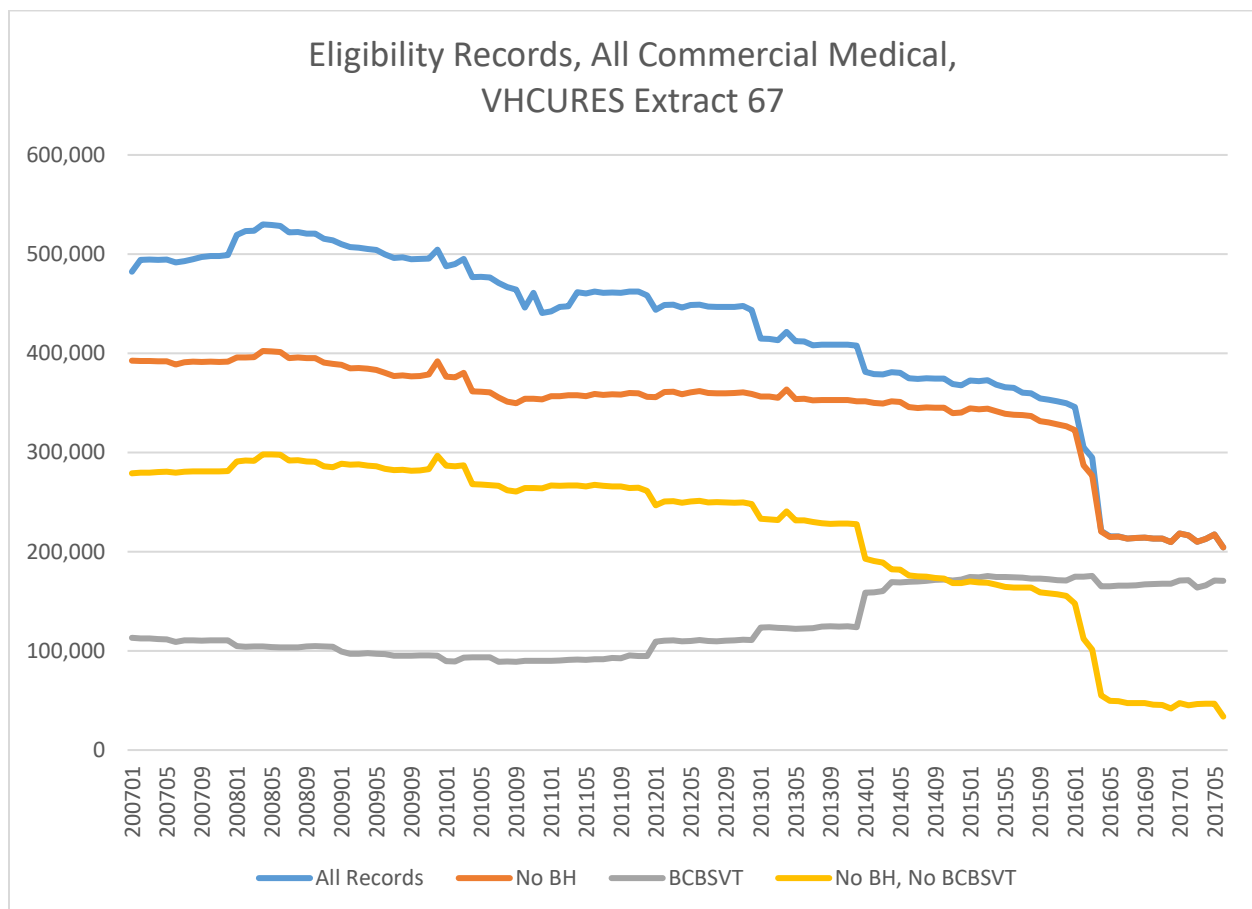
While a sizeable portion of the drop in non-BCBSVT commercial enrollment occurred in early 2016 (probably as a result of the March 2016 US Supreme Court decision in *Gobeille vs. Liberty Mutual*), enrollment was down by about 50 percent from 1/2008 to 12/2015.

¹⁴ Behavioral health plans are excluded because in many cases, the enrollment that they report is also reported by a medical plan.

Aggregate Analysis

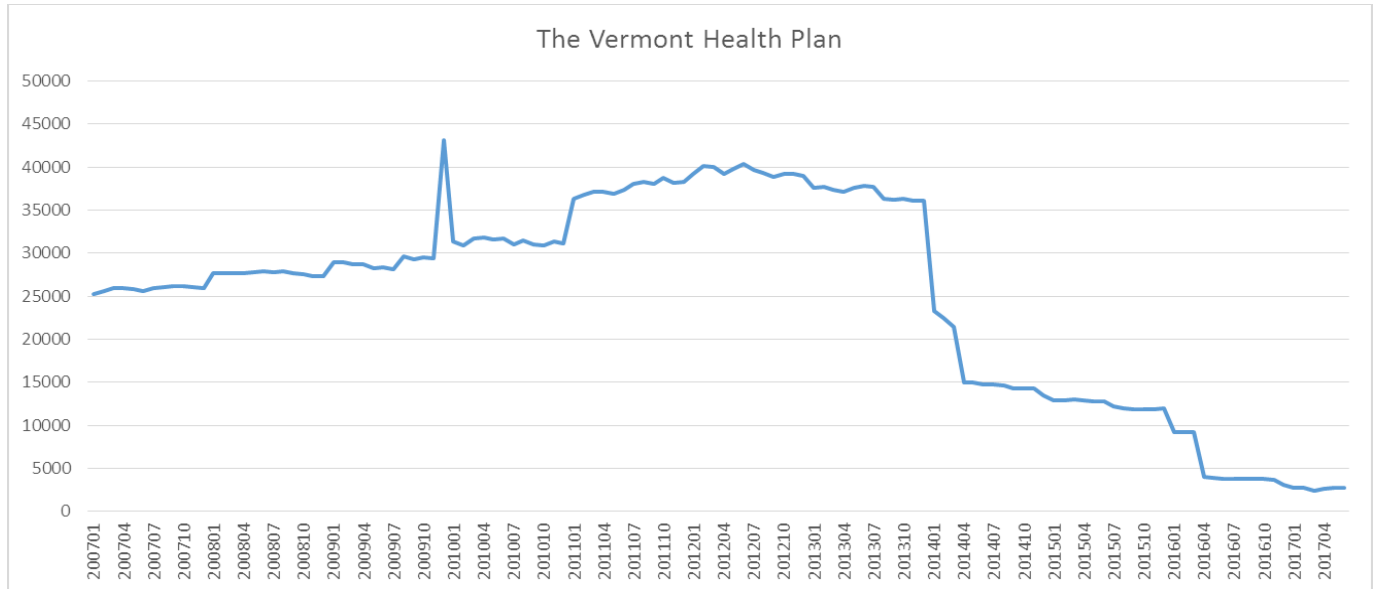
The chart below shows enrollment by month from January 2007 (the earliest data in VHCURES) through June 2017 (the most recent). As can be seen, enrollment in the aggregate declined slowly until early 2016, took a sharp drop in the first quarter of 2016, and then stabilized.

Within that pattern, an increase in BCBSVT enrollment offset the small decline in other commercial enrollment prior to 2016. This was primarily a shift from CIGNA to BCBSVT. However, the post-*Gobeille* decline in other commercial was not offset at all by BCBSVT.



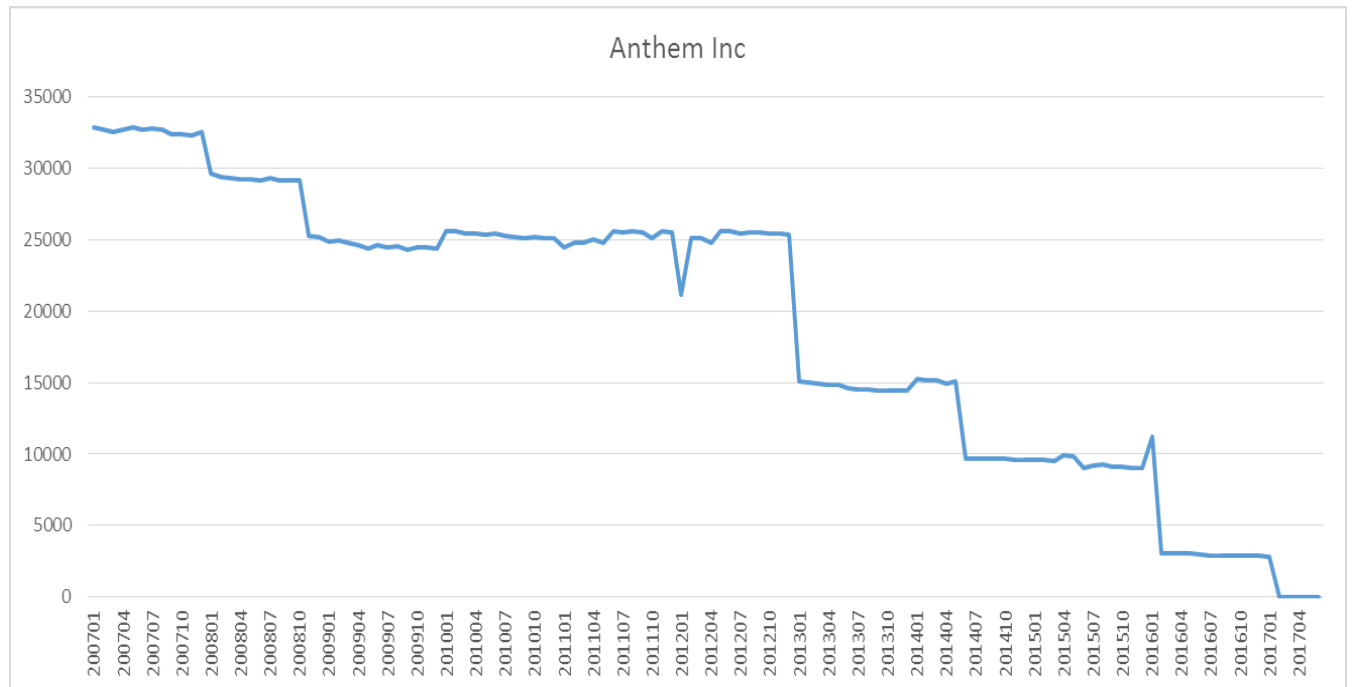
Selected Individual Plans

Looking at individual plans reveals different patterns. For example, The Vermont Health Plan¹⁵ (TVHP) showed growth until late 2013 (to about 40,000 lives), at which time enrollment dropped by about 60 percent. A further drop in early 2016 leaves the plan with about 3,000 lives.

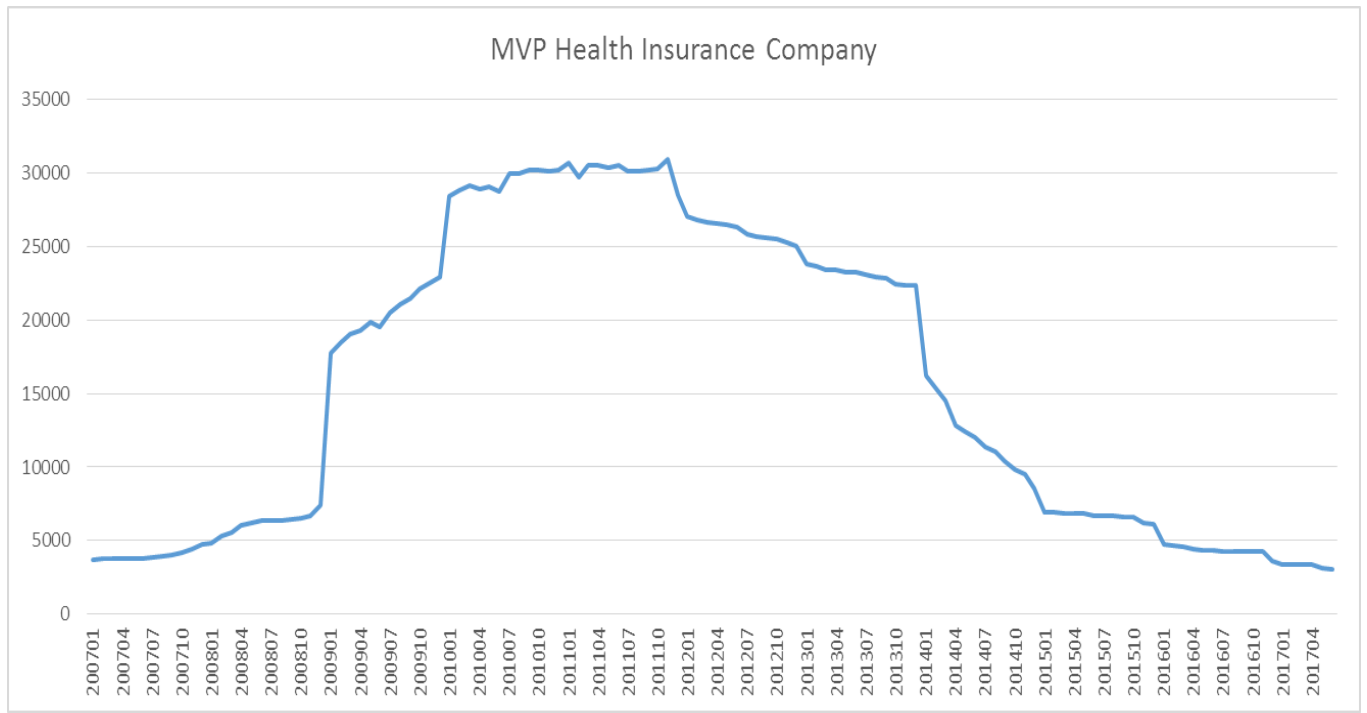


¹⁵ Although TVHP is a subsidiary of BCBSVT, it is reported separately in VHCURES and included in the non-BCBSVT category.

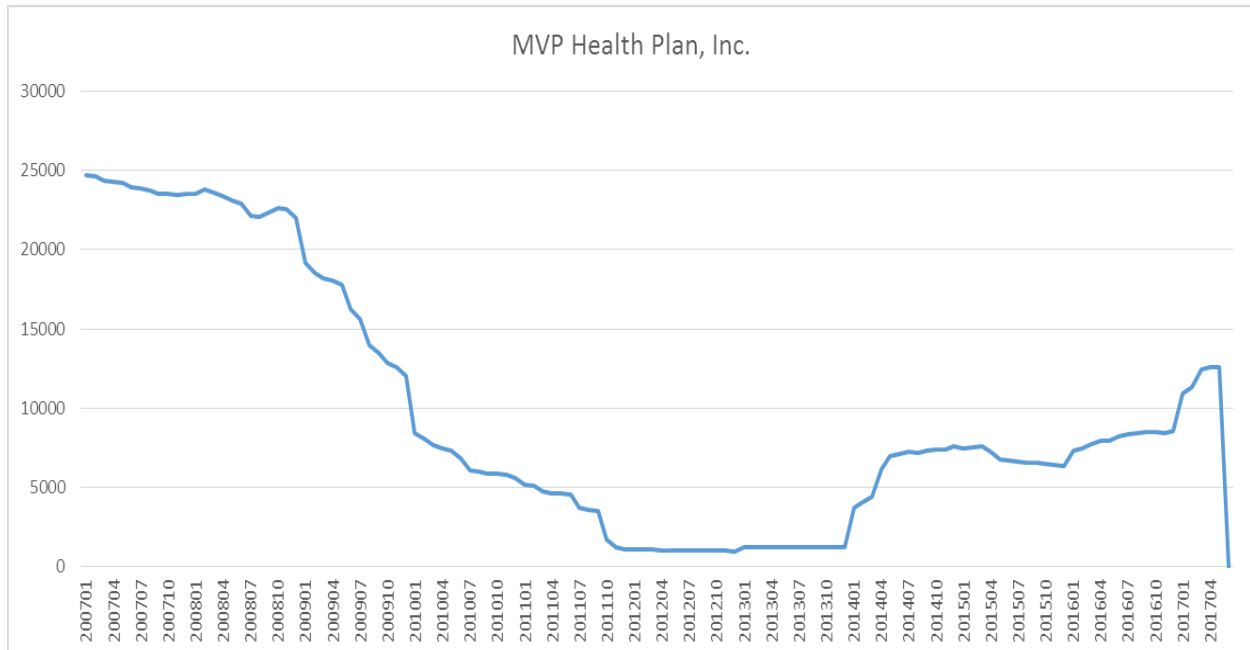
Anthem, which provides insurance in multiple states under the BlueCross trademark, shows multiple drops, falling from about 32,000 to about 2,800, and then to 0.



MVP Health Insurance, the non-HMO component of MVP grew from under 4,000 to about 30,000 from 2007 to mid-2010 and then fell back to 3,000 by June of 2017.



MVP Health Plan, the HMO business, showed a pattern that was somewhat an inversion of MVPHI, dropping from 25,000 to just over 1,000, rebounding to 12,600 before falling to zero in June 2017 (probably an issue in data submission).



Appendix 3 – Possible Data Model

This section offers a possible conceptual model that can be used for a stand-alone VHCURES data set, but could also support integration of claims and EHR data.

A useful metaphor for this model is beads on a string, where the string is an individual over time and the beads are encounters with the health care system. Each bead would contain the clinical information derived from that encounter, potentially including diagnoses, procedures, test results, and dates. Information on individual providers could be attached to the bead.

In addition to richer information (e.g., test results), the key difference between this model and a basic claims model is the treatment of diagnoses. A diagnosis is an attribute of a person and a period of time, NOT an encounter, although it can also be a reason for the encounter.

This difference is of particular importance when looking at the impact of chronic illnesses and creating episodes of care. A typical challenge in chronic disease epidemiology is the identification of people with diabetes. Currently, diabetes may or may not be reported on a claim for an unrelated condition. Often, researchers are forced to rely on medication data. In one interview the interviewee expressed a concern that there is no standard definition that can be used to identify people with diabetes so researchers must build their own.

Appendix 4 – Acronyms

ACC – Accountable Care Community
ACG – Adjusted Clinical Groups
ACH – Accountable Communities for Health
APCD - All-Payer Claims Database (e.g., VHCURES)
ACO – Accountable Care Organization
ADAP – Alcohol and Drug Abuse Programs
AHS – Agency of Human Services
AOA – Agency of Administration
APM – All-Payer Model
APMH – Advanced Practice Medical Home
BH – Behavioral Health
BHN – Behavioral Health Network
Blueprint - Blueprint for Health, Office of Health Access
BRFSS – Behavioral Risk Factor Surveillance System
CAGR – Cumulative Average Growth Rate
CAHPS – Consumer Assessment of Healthcare Providers and Systems
CBC – Complete Blood Count
CCIIO – The Center for Consumer Information & Insurance Oversight
CCMR – Care Coordination Medical Record
CHAC – Community Health Accountable Care, LLC
CHF – Congestive Heart Failure
CHIP – Children’s Health Insurance Program
CMMI – Center for Medicare and Medicaid Innovation
CMO – Chief Medical Officer
CMS – Centers for Medicare and Medicaid Services
COPD – Chronic Obstructive Pulmonary Disease
DAIL – Department of Disabilities, Aging, and Independent Living
DHMC – Dartmouth Hitchcock Medical Center
DID – Difference in differences
DRG - Diagnosis-Related Group
DUA – Data Use Agreement
DVHA – Department of Vermont Health Access
ED – Emergency Department
EHR – Electronic Health Record
EMR – Electronic Medical Record
EMT – Emergency Medical Technician
EOC – Episodes of Care
ePHI – Electronic Protected Health Information
ERG – Episode Risk Grouper
Esri - Environmental Systems Research Institute
ETL - Extract, Transform, and Load
FEDU – Frequent ED Use
FFS - Fee for Service
GIS - Geographic Information Systems
GMCB – Green Mountain Care Board

HC – Health Care
HCM – Health Confidence Measures
HDI – Health Data Infrastructure
HF – Healthfirst
HH – Health Home
HHS – U.S. Department of Health and Human Services
HIE – Health Information Exchange
HIPAA – Health Insurance Portability and Accountability Act
HIT – Health Information Technology
HITECH Health Information Technology for Economic and Clinical Health Act
HMO - Health Maintenance Organization
HP – Hospital Readmissions
HRQL – Health Related Quality of Life
HRR - Hospital Referral Region
HSA – Hospital Service Area (DOH)
HSA – Health Service Area (Blueprint for Health)
HSE – Health Services Enterprise
IBNR – Incurred But Not Reported
IDF - Identifiable Data Files
IFS – Integrated Family Services
IOM – Institute of Medicine
IT – Information Technology
LDS - Limited Data Set
LS – Learning Session
LTSS – Long-Term Services and Supports
MA – Medical Assistant
MEPS - Medical Expenditure Panel Survey
MD – Medical Doctor
MPI – Master Patient Index
NAACO – National Association of ACO's
NIST – National Institute of Standards and Technology
NPI – National Provider Identifier
NQF – National Quality Forum
OCR – The Office for Civil Rights within HHS
OCV – OneCare Vermont
ONC – The Office of the National Coordinator for HIT w/in HHS
OS – Operating System
P4P – Pay for Performance
PCMH – Patient Centered Medical Home
PCP – Primary Care Physician
PDF – Portable Document Format
PHI – Protected Health Information
PII - Personal Identifiable Information
PPS – Prospective Payment System
PRG – Pharmacy Risk Grouper
PUF - Public Use File
QCCM – Quality and Care Coordination Manager
QI – Quality Improvement

RFP – Request for Proposal
RN – Registered Nurse
RUI – Resource Use Index
SAS – Statistical Analysis System
SBIRT – Screening, Brief Intervention, and Referral to Treatment
SCÜP – Shared Care Plan/Universal Transfer Protocol
SMHP - State Medicaid Health Information Technology Plan
SMS – Short Message Service
SOV – State of Vermont
SPA – State Plan Amendment
SPC – Statistical Process Control
SRA Tool – Security Risk Assessment Tool
TACO – Totally Accountable Care Organization
TBD – To be determined
TCI – Total Cost Index
TCM – Transitional Care Model
TCN – Transitional Care Nurse
TCOC – Total Cost of Care
TCRRV – Total Care Relative Resource Value
UVMCM-University of Vermont Medical Center
VCGI – Vermont Center for Geographic Information
VCN – Vermont Care Network
VCP – Vermont Care Partners
VCP – Vermont Collaborative Physicians
VDH – Vermont Department of Health
VHCIP – Vermont Health Care Innovation Project
VHCURES – Vermont Healthcare Claims Uniform Reporting & Evaluation System
VHIE – Vermont’s Health Information Exchange
VITL – Vermont Information Technology Leaders
VPQHC – Vermont Program for Quality in Health Care
VT – Vermont
VUHDDS -Vermont Uniform Hospital Discharge Dataset